Polysubstance Use Patterns among Women with at-Risk Alcohol Consumption

by

Jian Shi

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science

Baltimore, Maryland
April, 2018

Abstract

Introduction

The complex pattern of polysubstance use poses methodological challenges to research in both the measurement and statistical analysis. The common strategy of data collection in prior studies, the quantity-frequency method, has a limitation in capturing the concurrency of multiple drug use. This study looked for an alternative method to more accurately identify patterns of polysubstance use, therefore provide a more precise measure of exposure for studying the prevalence of polysubstance use and its relationship with public health burdens.

Methods

We used baseline data from a randomized controlled trial testing a brief alcohol reduction intervention in 2 urban Sexually Transmitted Infections (STI) clinics. The study included 439 women between 18 and 66 years old, who had met criteria for at-risk alcohol use (more than 7 standard drinks per day or >=4 drinks on one occasions or reported sex under the influence of alcohol) in the prior three months. Daily use of alcohol, marijuana, cocaine, heroin, and prescription drugs were collected using a 30-day Time Line Follow Back (TLFB) interview . Demographic characteristics were collected at baseline. Hierarchical clustering with unsupervised random forest was used for clustering the polysubstance use patterns on the one-day, two-day, and weekly level. Multistate model was used to investigate changes of substance use patterns and to examine the association between nicotine addiction and polysubstance use pattern transitions.

Results

On the day level, patterns of substance use were categorized into 10 groups, namely, 1) abstinence, 2) mild to moderate alcohol use, 3) binge alcohol use, 4) marijuana, 5) marijuana and mild to moderate alcohol use, 6) binge drinking and some marijuana use, 7) binge drinking and marijuana use, 8) cocaine and heroin use, 9) mild to moderate alcohol use and some use of marijuana, cocaine, heroin, and prescription drugs, 10) binge drinking, and some use of marijuana, cocaine, heroin, and prescription drugs. The abstinence was the most frequent pattern with the longest sojourn time of 3.89 days. On the weekly level, the most frequent weekly pattern was having a positive proportion of binge alcohol, marijuana, cocaine (<0.2), heroin (<0.05), and prescription drug (<0.05) use on each day of the week with an increase use in each substance, except prescription drugs, on Friday and Saturday. It had the longest sojourn time of 4.33 weeks.

Conclusions

Awareness of the concurrency of polysubstance use should be raised in the measurement of substance patterns and the interpretation of study results. TLFB and unsupervised random forest provide an approach for a more accurate categorization of polysubstance use.

First Reader: Bryan Lau

Second Reader: Geetanjali Chander

Acknowledgements

I would like to acknowledge my deepest thanks and gratitude to Dr. Bryan Lau, for suggesting the topic, and his patience and unfailing support during the study. His aspiring guidance in addressing the methodological challenges and thesis writing has been essential during this work and invaluable inspiration for my future research.

I am also gratefully indebted to my other advisor, Dr. John Jackson, for generously sharing his recent work, being greatly helpful during my search for potential topics, and providing insightful discussions about the research.

I would like to express my warmest thanks and sincere appreciation to Dr. Geentanjali, for her generous support of data sharing. Her encouragement and constructive criticism also helped me to develop a broader perspective of my research and convey a clearer message in the thesis.

Last but not least, special thanks to Dr. Corrine Joshu for giving advice on thesis writing in the monthly meeting, and Frances Burman and Laura Camarata for coordinating the project.

Contents

Introduction

Substance use refers to the consumption of psychoactive substances, including illicit drugs, alcohol, and tobacco, over a specific period of time. The prevalence of substance use in the United States has remained high in the past few years. In 2016, 28.6 million (10.6%) Americans aged 12 or older, were current illicit drug users. An estimated 65.3 million (24.2%) Americans aged 12 years or older, reported binge drinking (≥5 drinks for males and ≥4 drinks for females on the same occasion on at least 1 day) in the past month[1]. Substance use poses serious public health concerns in the United States. It can lead to short- and long- term health consequences both mentally and physically, including addiction, dependence symptoms, depression, cancer, and infectious disease [2]. Apart from the direct health impacts on the people who take the substance, the people around them in a family or a community can also be affected in various aspects, e.g., relationship [3], and criminal involvement [4]. It is estimated that the annual cost for Americans to address the impact of substance use is over $600 billion [5].

Single drug use has become progressively scarce in the community and clinical settings [6]. Multiple drugs can be taken simultaneously in order to strengthen the effect on central nervous system [7] and create an intense high, or sequentially to prevent withdrawal symptoms [8] or simply to take others as alternatives when the one that is dependent on is not available. The term "Polysubstance use" broadly describes consumption of more than one substance within a defined period. It is usually considered as more hazardous than single substance use, because it can boost toxicity, and cause overdose and mental health disorders [9, 10, 11, 12]. Due to the possible interactions of different drugs, the effect of polysubstance use is not equal to the sum of the effect of each substance used separately. Hence, prior studies on one single type of substances render limited implications for polysubstance use, and direct investigation in multiple drug use is needed.

In fact, considerable studies have been conducted examining the risk factors [13,14] and health consequences [15,16] of polysubstance use, as well as the efficacy of interventions[17]. However, identification of polysubstance use profiles has always been a challenge owing to the practical obstacle in collecting every combinations of drugs used in the past and their frequencies, as well as the methodological challenge to appropriately classify them for analysis when the detailed information is available. Polysubstance patterns were categorized based on the number of substance types in some studies, and on the different combinations of substance types in others. Among the studies with the latter criteria for categorization, the length of periods the substance use was measured for varied from two weeks to lifetime [18], yet none of the resulted substance use profiles managed to exclude the potential heterogeneity between concomitant (taken at the same time) and sequential (one drug followed by another) multiple substance use. Accurately speaking, polysubstance use was defined as having a multiple-drug-use history during the specified period in these studies. The results of the studies were thus the overall estimated effect of multiple drug use on the health consequences, averaging out the potential variation between the concurrent use and sequential use. Both risk factors and health consequences may both differ between the two behaviors, since the motivation of drug use and the level of drug interactions can be different. Therefore, the variation should be aware of when interpreting the results and utilizing that for guiding the intervention.

It would be scientifically meaningful if more accurate identification of polysubstance use patterns can be developed so that the prevalence of polysubstance use and its relationship to public health burden can be measured with a better defined exposure. To accomplish that, there are mainly two challenges to be addressed, namely accurate measures of polysubstance use and appropriate statistical analysis methods. Timeline follow back (TLFB), a retrospective calendar-based, self-reported measure of daily substance use, has been widely employed in research for use of both alcohol, which was initially designed for, and has also been extended to marijuana and other illicit drugs [2, 19]. TLFB has been shown to have a high agreement with biological measures for substance use [20], and considered as a psychometrically sound instrument for the accuracy of the measurement and the sensitivity to changes in substance use [19, 21]. With more specific information collected using TLFB, the data dimensionality is further increased compared with the commonly used quantity-frequency (QF) measurement. Though the data dimensionality was relatively low, previous studies using QF still faced impediment to statistical analysis due to small cell frequencies [18]. Latent class analysis (LCA) was the method most frequently used to deal with the high dimensionality of data in recent studies [18]. It derives classes with probabilistic model describing the distribution of the data and thus avoids bias that may exist in the traditional clustering approaches due to arbitrarily chosen measure of dissimilarity [22]. However, LCA has its limitation dealing with continuous indicators (needed to be discretized). Unsupervised random forest, although is new to field of polysubstance use, has been successfully used in genomic analysis for clustering [23,24], and can be considered as an alternative to address the issue.

The current study aimed to bridge the gap by exploring the daily substance use pattern, as well as every 2 days and weekly during the 30 days of period, and assessing the stability of the pattern for each individual across time. The study leveraged use of a 30-day timeline follow back (TLFB) data of women with at-risk alcohol use attending the urban STI clinic. The use of multiple drugs in the same day was approximated as concurrent. Hierarchical clustering with unsupervised random forest was employed to develop clusters of polysubstance use patterns. The transition among different polysubstance use patterns were explored utilizing the 2-day and weekly clusters, and multistate model to estimate the transition probability and sojourn times based on the daily clusters.

Methods

Subjects

Participants were recruited between September 2012 and May 2015 from two STI clinics in Baltimore, MD. Participants were enrolled for a randomized clinical trial for computerized brief alcohol intervention as a part of the "SHARP Women Study". We restricted the data to the investigation of their substance use patterns before intervention. Women were eligible for enrollment of the trial if they were ≥18 years of age, not pregnant, drank at risky levels (defined as >7 standard drinks per day or >=4 drinks on one occasions or reported sex under the influence of alcohol) in the prior three months , had their own cell phone and used text messaging (for later intervention), and would not move out in the next 12 months.. All 439 women enrolled in the trial were included in this study.

Measures

Daily use of substances, including alcohol, marijuana, cocaine, heroin, and prescription drugs was measured with the TLFB [19, 25] method over the 30-day period prior to baseline. Among the five types of substances, alcohol use was collected as the number of standard drinks the individual had on that day, other four drugs were collected as binary variable [25]. From the clinical perspective, alcohol use was also classified as binge or not, with the cutoff of four or more standard drinks for women, approximate to blood alcohol concentration of 0.08g/dL defined as binge by National Institute of Alcohol Abuse and Alcoholism (HIAAA) [26]. Data of 13167 days were recorded with 1-day missing data from 2 participants respectively. We excluded 77 days from 16 participants due to controlled environment to avoid their potential effect on substance use change. The controlled environment included hospital, incarceration, inpatient substance abuse rehabilitation program, nursing home and group home. Age, race, education, marital status, children, living situation, employment, income, smoking, and HIV infection were collected at Baseline.

Statistical Analysis

Polysubstance use patterns were investigated on the one-day, two-day, and week bases. The two-day-level data were generated from the 30 days for each individual by a two-day moving window. In order to capture potential regularity of polysubstance use pattern on certain weekdays and weekends, the week-level data included any whole weeks from Saturday to

Sunday for each individual, instead of being generated by a seven-day moving window. To deal with the high dimensionality of the data, unsupervised random forest was conducted to generate the dissimilarity measure, with which hierarchical clustering was employed to form clusters for polysubstance use patterns on the three level bases respectively.

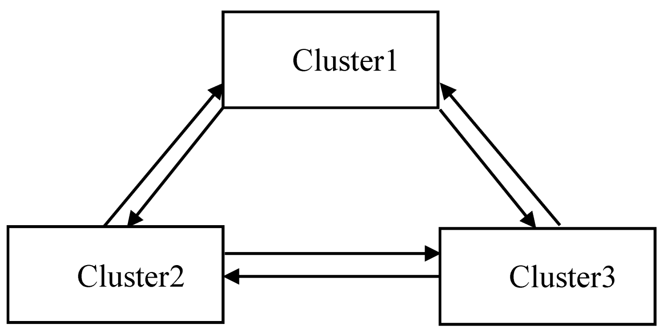
Random forest is an ensemble statistical learning method for both classification and regression [27] (in this study, it was used for classification). The main steps involved in the procedure include 1) generating "pseudo training sets" via bootstrap, 2) growing trees (training the method) on each training sets, with a subset of the predictors randomly sampled as candidates for each splitting node, 3) averaging all the predictions from the trees by the majority vote. Bootstrap aggregation (bagging) averages all the predictions of bootstrapped samples from the original dataset, which reduces the variance due to overfitting of each tree while has the advantage over pruning classification/ regression trees in that bagging trees can grow deep so that the prediction bias does not increase [28]. Random forest is a further improvement of bagged trees, which de-correlates trees by randomly selecting candidate variables for each splitting so that trees can grow deeper to reduce the prediction bias [28]. This was especially relevant in this study due the high correlation between binge drinking and drinking quantity. While random forest is mainly employed for supervised learning, it can also be used to produce in unsupervised learning to produce proximity matrix, which can then be transformed to dissimilarity measure for clustering (1). This method has been successfully used in some genomic analysis [23, 24]. The main idea is to suitably generate synthetic data labeled as the artificial class, and construct random forest predictors to distinguish the synthetic data and observed data [23]. Compared with traditional dissimilarity measure, e.g., Euclidean distance, unsupervised random forest has the advantage of being robust to outliers and handling mixed types of variables well [23], which is helpful in dealing with one continuous variables, i.e., drinking quantity, together with other five binary variables. *Random Forest SRC* package in R was used for the unsupervised random forest. The number of variables randomly selected as candidates for splitting, *mtry*, was set to the default, which is one third of the total number of predictors. The number of trees, ntree, was also set to the default of 500.

$$\text{Dissimilarity}=\sqrt{1-Proximity} \quad (1)$$

Hierarchical clustering is a widely applied approach to detect clusters in many different areas, ranging from biology to genomics to computer vision [29, 30, 31]. It is a bottom-up method, which generates the dendrogram by starting at each object within its own cluster, and identifying the closest object to merge into a new cluster until all the objects were in one cluster [28]. Not like K-means clustering, it does not require a pre-specified number of clusters, K [28]. *Hclust* function in R was used for hierarchical clustering. The *Dynamic Hybrid* method, available in the R package *of Dynamic Tree Cut*, was used for detect clusters in the dendrogram, which is good at outlier detecting and outperformed the Static height cutoff method in the gene data analysis by Langfeler et al [32]. Given low frequencies in some clusters, the clusters that were very close, always only moderately different in drinking quantity, were merged for further analysis on the one-day level. For the convenience of visual display of transitions, clusters were recoded with the same types of substances coded close to each other, and in the order of the potential risk of hazardous effects in large.

The substance use patterns for each individual during the 30 days were shown in the lasagna plot, which is a heatmap adapted to display the longitudinal data [33]. The stability of polysubstance use for each individual across time was assessed through the trend of polysubstance use pattern in the two-day and week base clusters. It was also examined via multistate model with the one-day base clusters by estimating the transition probability from one cluster to another and sojourn time of each cluster. The sojourn time refers to the expected time the person spends in one state (cluster) before leaving for another. model is often seen as an extension to competing risk model, which deals with intermediate states [34], and does not require one single transition and several absorbing final states. Figure1 shows one example of the model. Any of the three clusters can be the initial state for individuals, and transition between any two of them is possible. The p_{12} in the transition probability matrix represents the probability of transitioning from Cluster1 to Cluster2. The *msm* package in R was used for the study. The results of the multistate model were stored as transition intensities, the instantaneous risk of moving from state r to s, from which the transition probability can be calculated. The Markov model assumes that the next state the individual will be in and the time at which this will happen only depend on the present state. With days of controlled environment excluded, this assumption was likely to hold for the study. As the day was taken as the time unit, and the substance use on each day was assumed to occur concurrently, the observed transition times were considered as the exact times. The effect of age and smoking addiction on the transition were also examined in the model. Due to the limitation of the number of variables that can be included in the model given some small cell frequencies of certain clusters, smoking addiction was only estimated with the time to the first cigarette after waking up, and was reclassified as “don’t smoke”, “over 60 minutes”, and “within 60 minutes”, and age was taken as a continuous variable. The effect of the covariates on transition was assessed through the hazard ratio. The transition among clusters of polysubstance use on the week-level was also examined.

The results of the clustering on the day level, two-day level, and week level, were visually displayed in the spider plot made in Microsoft Excel (2016). The statistical learning, i.e., random forest, hierarchical clustering, and dynamic tree cut, was done in R (version3.4.0). Other statistical analysis, including the spaghetti plot of Figure4. was done in STATA 14.



$$P=\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

Figure1.

Results

All the 439 women enrolled in the randomized clinical trial of brief alcohol intervention were included in the analysis for polysubstance use patterns in 30 days before intervention. The demographics of the participants were shown in Table1. The participants were predominantly Black women except for one. The median age was 31 years old, with the range from 18 to 66. Of the 439 women, 72.0% had never married, and 68.1% had children. 52.8% of the participants had their own apartment, 82.2% were full-time employed, and 82.5% had an annual income equal or less than $15,000. Current smokers accounted for 56.0% of the participants, 85.8% of whom had the first morning cigarette within 60 minutes after waking up. A small proportion of participants had HIV (8.7%).

Determining concurrent polysubstance use by reporting the number of substances used in the past 30 days does not capture the heterogeneity in the use of multiple substances. Table 2 showed the number of substances reported being used in the past 30 days, whereas data from the TLFB provides the median proportion of days reporting the concurrent use.

Participants who reported to have used single substance had a median proportion of 0.17 of the days taking 1 substance during the 30-day period. Participants who reported to have used 2 types of substances had a median proportion of 0.07 of the days taking 2 substances during the 30-day period. Participants who reported to have used 3 types of substances, had a median proportion of 0.03 of the days taking 3 substances during the 30-day period. Participants who reported to have used 4 types of substances, had an expected proportion of zero of the days taking 4 substances during the 30-day period. Participants who reported to have used 5 types of substances, had an expected proportion of 0.03 of the days taking 5 substances during the 30-day period.

Polysubstance use pattern on the one-day level

After Excluding the 77 days of controlled environment and 2 days of missing records, a total of 13090 days from 439 participants were included in the analysis of the polysubstance use on the day level. We used hierarchical clustering with unsupervised random forest to form clusters for the daily polysubstance use patterns. Thirteen clusters were finally generated with the Dynamic Tree Cut. (Table3, Figure2.)

Cluster1 had the largest sample size, 49.36% of the day observations. It comprised the days when individuals did not take any of the five types of substances. Cluster5, Cluster10, and Cluster13, consisting of 4.19%, 1.12%, and 0.84% of the observations, referred to the days at low risk of negative health consequences, when individuals only had low quantity of drinks with the average of 2.22, 1.14, 0.75 standard drinks respectively without any other drugs combined. Cluster2, Cluster3, and Cluster8 were composed of single substance use, either binge alcohol use or marijuana. Cluster2, consisting of 13.14% of the observations, referred to the days when the individual took marijuana without any of other four substances combined. Cluster3 comprised 10.18% of the observations. Individuals on the day of Cluster3 had binge drinking, with an average of 5.59 standard drinks without any other drugs combined. Cluster8 consisted of 3.51% of the observations. Individuals had binge drinking with an average of 11.12 standard drinks without any other drugs combined. Cluster4, Cluster6, and Cluster9 had use of both alcohol and marijuana. Cluster4 consisted of 6.05% of the observations. Individuals on the day of Cluster4 had binge drinking, with the average of 7.16 drinks and all took marijuana. Cluster6 comprised 3.73% of the observations, with an average of 23.33 drinks and 39% marijuana use. Cluster9 consisted of 2.48% of the observations. It referred to the days when individuals had an average of 1.84 standard drinks without binge, and all had marijuana use. Cluster 11 referred to 60% of the observations using cocaine and 40% using heroin, counting for 1.00% of the observations. Cluster7 and Cluster12 were two clusters likely to have the highest risk of hazardous effects. Cluster 7 comprised 3.54% of the observations. The proportion of reported substance use on the day of the cluster was positive for all five types of substances, namely, 0.98 for binge, 0.24 for marijuana, 0.61 for cocaine, 0.37 for heroin, and 0.17 for prescription drugs. The mean of drinks individuals had on the day of this cluster was 14.33. Cluster12 comprised 0.86% of the observations. No binge drinking was reported on the day of the cluster, with an average of 0.77 standard drinking, but all the four drugs were reported for use with proportions of 0.38 for marijuana, 0.42 for cocaine, 0.30 for heroin, and 0.38 for prescription drugs.

Transition among different substance use patterns on the day level

Participants had been in four different clusters during the 30 days on average. In order to further investigate the transition among clusters for individuals, transition probabilities and sojourn times were calculated using multistate model. The effect of age and time to first cigarette after waking up (as an estimate of nicotine addiction) on the transition were examined. Given the small frequencies of some clusters, clusters that were similar in terms of the types of substances used and potential risk of hazardous effect, namely, Cluster5, Cluster10, and Cluster13 as having low quantity of drinks only, as well as Cluster3 and Cluster8 as binge drinking only, were merged for the analysis. The clusters were regrouped based on the description of patterns noted in Table4.

The substance use pattern during the 30 days of the 439 participants were shown in the lasagna plot. Each row represented one individual, and each column corresponded to the day. The rows in the plot was sorted by the cluster on the first day, with the gradient of color corresponding to the number of patterns (lightest, 1; darkest, 10; white, missing data/excluded for controlled environment) (Figure3). In general, people tended to alternate among similar clusters (with similar color), and abstinence was commonly taken as an interim for all the groups based on the first day substance use. There was a large proportion of the study population did not use any substance for more than half of the month, while took alcohol or marijuana, or the two combined occasionally. There was also a few proportion with single substance use of marijuana as a predominant substance use pattern with occasional binge drinking. Some individuals used cocaine, heroin, and prescription drugs occasionally, while some took those with or without marijuana or alcohol for most of the days.

The probabilities for each transition between two patterns in consecutive two days were listed in Table3. The column represented the current state and the row represented the state in the next day. The values on the diagonal were the probabilities remaining in the same state for two consecutive days, which all turned out to be the largest among their respective row, indicating that regardless of the current state, participants tended to remain in the same cluster for the next day. This was consistent with the sojourn times (Table5), which were all above 1 day. Among the 10 states, abstinence had the longest sojourn time of 3.89 days, meaning that people on average remained as not having any substance use for 3.89 days. The next two states following in the rank were mild to moderate alcohol use with some use of marijuana, cocaine, heroin, and prescription drugs, and single substance use of marijuana, with a respective sojourn time of 2.69 days and 2.57 days, suggesting that these two patterns of substance use were more likely to maintain in the next day compared with other substance use pattern (except for abstinence). The state of binge drinking with some use of marijuana, cocaine, heroin, and prescription drugs was considered with the highest risk of negative health consequences. The polysubstance use patterns across time for this pattern during the 30 days were shown in Figure4. The long sojourn time may be attributed to the individuals who predominantly stayed in this highest risky pattern during the 30 days, such as subject 25, 69, 119, and 185. Some individuals remained in the last three patterns for most of time, e.g., subject 161, 245, and 254, suggesting their frequent use of cocaine, heroin, and prescription drugs. There were also quite a few individuals who predominantly stayed in low risky patterns, e.g., abstinence and mild to moderate drinking, and jumped into the highest risky pattern occasionally.

Among the days when substance use patterns changed on the next day, the probabilities of transitioning to each other patterns were shown in Table6. Abstinence had a positive probability of transitioning to any other patterns. Binge drinking and mild to moderate alcohol use were the two most frequent patterns following abstinence, with a respective probability of 0.43 and 0.24. This high probability of transition was also mutual, as the probabilities of transitioning from binge drinking to abstinence and from mild to moderate alcohol use to abstinence were 0.72 and 0.69, the highest among all the clusters for the probabilities of transitioning to abstinence. There were a total of 72 individuals among the 439 that had the substance use patter alternating among abstinence, mild to moderate alcohol use, and binge drinking during the 30 days, 19 individuals between abstinence and mild to moderate alcohol use, and 26 individuals between abstinence and binge drinking. That accounted for about one-fourth participants that only used alcohol without any of other four substances. Single substance use of marijuana had a probability of 0.33 transitioning to abstinence, 0.23 to marijuana use with mild to moderate alcohol use, 0.26 to binge drinking and marijuana use. The two states of marijuana use with mild to moderate alcohol use, and binge drinking with marijuana use also had a higher probability of transitioning to single substance use of marijuana compared with transitioning to other states, with a respective probability of 0.49, and 0.39. These three patterns all had 100% taking marijuana, only with a difference in alcohol intake, namely 0 for pattern 4, light drinking for pattern5, and binge drinking for pattern 7. This described an important transition pattern for some individuals predominantly taking marijuana, where they remained taking marijuana with alcohol use fluctuating between days. The other similar cluster in terms of types of substances was binge and some marijuana use with a median alcohol use of 19 standard drinks, as well as 0.39 proportions of taking marijuana, which mainly transitioned to binge or (and) marijuana use. Some cocaine and heroin use had a 0.60 probability of transitioning to abstinence, and 0.21 probability transitioning to the state of binge drinking with some use of marijuana, cocaine, heroin, and prescription drugs. It tended to transition to binge compared with mild to moderate drinking, and transition to concurrent use of marijuana and other drugs than marijuana use only. The pattern of mild to moderate drinking with some use of marijuana, cocaine, heroin, and prescription drugs had a dispersed distribution of the next state, with a probability of 0.13 transitioning to abstinence, 0.08 to mild to moderate drinking, 0.15 to binge, 0.14 to single substance use of marijuana, 0.10 to binge and some marijuana use, and 0.09 to binge drinking with some use of marijuana, cocaine, heroin, and prescription drugs. The pattern of binge drinking with some use of marijuana, cocaine, heroin, and prescription drugs had nearly half chance of transitioning to abstinence, a probability of 0.12 transitioning to binge only, and a probability of 0.13 to mild to moderate drinking with some use of marijuana, cocaine, heroin, and prescription drugs.

The effect of covariates was assessed based on the covariate estimate (Table7). While limited information (large 95% confidence intervals) was provided owing to the small frequencies of some clusters, there appeared to be an association between having the first cigarette within 60 minutes and the transition in abstinence. The individuals who had the first cigarette within 60 minutes had a 0.37 (95% confidence interval: 0.22, 0.50) lower hazard ratio of transitioning from abstinence to mild to moderate drinking over maintaining abstinence compared to non-smokers. However, they had a statistically significant higher hazard ratio of transitioning from abstinence to binge, marijuana, or combinations of the five substances over maintaining abstinence compared with non-smokers.

The fluctuations of polysubstance use patterns between two consecutive days were displayed by clustering on the two-day level (Figure5). A total of 12,637 observations of polysubstance use were generated with a 2-day moving time window based on the 10,390 observations on the day level. Seventeen clusters were then generated with the same method as the daily level. 35.55% of the observations were no-substance use for both days. Apart from the no-substance use cluster, the overall shape of the cluster can be categorized into two groups. One was that the two days mainly differed in whether one (or more) type(s) of substance was used. For instance, Cluster5 had binge on the first day while not on the second day. This group contained 24.89% of the observations, including Cluster5, 6, 7, 8, 9, 11, 14, 16, and 17. The other group was that the substance use was nearly symmetric between the two days. For instance, Cluster 4 only had marijuana on both days. This group contained 35.55% of the observations, including Cluster2, 3, 4, 10, 12, 13, and 15.

Polysubstance use pattern on the week level

A total of 15 clusters were generated from 1470 week observations on the week level (Figure6). All the clusters, except the abstinence (Cluster2) and Cluster4, had an evident surge of certain substance use on Friday or weekend. The increase of use on Friday or weekend did not only occurred to alcohol but also other four drugs. The most frequent weekly pattern was Cluster1, which had a positive proportion of binge, marijuana, cocaine (<0.2), heroin (<0.05), prescription drug (<0.05) on each day of the week with an increase use in each substance, except prescription drugs, on Friday and Saturday. This was different from the day level, where no-substance use was the most common one. The use of substances in Cluster4 fluctuated during the week with a slight increase in heroin and alcohol on Monday and decrease in alcohol use on Friday and Saturday.

Transition among clusters on the week level

The probability of transitioning to each cluster was displayed in Table8, and the mean sojourn time of the week level was displayed in Table9. Unlike the frequency distribution of clusters the day level, the cluster that had no use of any substance, i.e., Cluster2, was not the one that had the longest sojourn time, which can be attributed to the enrollment criteria of the study that each participant had at-risk alcohol use. Consistent with largest proportion of observations, Cluster1 had the longest sojourn time of 4.33 weeks, which suggested that individuals who had a week polysubstance use pattern of Cluster1 were expected to remain in the same pattern for the whole month. Cluster1 was also the cluster to which many clusters transitioned to.

Discussion

The study leveraged on the richness of data with a relatively small sample size to examine specific day-to-day polysubstance use patterns. On the day level, there were mainly three groups of individuals in terms of the transition: 1) alcohol use only, alternating among no-substance use, light to moderate drinking, and binge alcohol use, 2) marijuana and alcohol use, keeping marijuana use for most of the time while drinking or binge drinking occasionally, 3) cocaine and heroin use, with some having cocaine, heroin, prescription drugs, marijuana, and binge drinking for most of the days, and some alternating between the taking the five substances, and abstinence or binge drinking or (and) marijuana use. For the substance use in the two-day moving window, apart from the 35.55% observations did not use any substance on both days, 39.55% observations had a symmetric substance use pattern between the two days, and 23.22% were not symmetric. Among the non-symmetric clusters, the difference between the two consecutive days mostly lay in the alcohol use and a few in marijuana use. However, the detection of the evident changes in alcohol and marijuana use may not be a simple reflect of the properties of the two substances, but partly attributed to the overwhelming proportion of users of the two substances and alcohol was collected as the continuous variable, yet a low proportion of users of cocaine, heroin, and prescription drugs in the study population. Based on the weekly pattern of polysubstance use, an increase of use on Friday and weekend was observed in all five substances respectively or with some combined. Cluster1 on the week level, where there was a positive proportion for all the five substance use on each day of the week with an increase use in all substances except prescription drugs on Friday and Saturday, was estimated to remain in the same cluster during all the 30 days, indicating that there was a group of individuals in the study population who were highly dependent on multiple drugs, including both soft and hard drugs. Compared with the large proportion of abstinence on the day level, it suggested that the study population had a low proportion (9.05% on average) of having the whole week of no-substance use, and the day of abstinence mainly served as an interim between two days of some substance use. There was also some evidence suggesting the association between nicotine addiction, estimated by the time to first cigarette after waking up, and the transition of polysubstance use patterns. A crude assessment indicated that the common strategy to measure polysubstance use by asking the types of substances used during the last 30 days represented less than 8 days of concurrent polysubstance use on average. Therefore, the effect of multiple drug use assessed based on that measurement of polysubstance use may not represent the effect of the concomitant polysubstance use pattern.

The study has a few limitations. First, the study population were predominantly Black, non-Hispanic adult women attending the STD clinic, 82.2% of whom were full-time employed, and 82.5% of whom had annual income equal to or below $15,000. Therefore, the generalizability in terms of the prevalence of polysubstance use and clustering results of polysubstance use was limited. Second, dose of marijuana, cocaine, heroin, and prescription drug use was not measured, thus, changes of these drugs were not captured as well as alcohol. Finally, the frequencies of cocaine, heroin, and prescription drugs were very low in the participants, posing the difficulties in running statistical analysis.

The study has important implications for future research and interventions in polysubstance use in the specific social and cultural setting, and provides an application of machine learning to descriptive epidemiology for data mining. Awareness of the concurrency of polysubstance use should be raised in the measurement of substance patterns and the interpretation of study results. Compared with the common assessment of types and frequencies for polysubstance use, the TLFB method has an exclusive advantage of improving the accuracy of identification of polysubstance use patterns, which provides measures of time-varying exposures to examine short-term effects of polysubstance use. It provides guidance for interventions by helping to better understand polysubstance use patterns in a specific population, and to evaluate the effect of intervention on particular substance, e.g., alcohol, as the reduction of daily intake can be calculated. Further research can be done based on this study to investigate the association between specific polysubstance use patterns and short- and long-term health consequences, as well as the effect of alcohol intervention among individuals with different polysubstance use patterns.

Bibliography

1. Substance Abuse and Mental Health Services Administration. *Key substance use and mental health indicators in the United States: Results from the 2016 National Survey on Drug Use and Health*. In: HHS. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/ [Accessed 2 March 2018].

2. Connor J, Gullo M, White A, Kelly A. Polysubstance use: Diagnostic challenges, patterns of use and health. *Current Opinion in Psychiatry*. 2014;27(4):269-275. http://www.ncbi.nlm.nih.gov/pubmed/24852056. doi: 10.1097/YCO.0000000000000069.

3. Center for Substance Abuse Treatment (CSAT). Chapter 2 Impact of Substance Abuse on Families. In: SAMHSA/CSAT treatment improvement protocols. *Substance abuse treatment and family therapy.* Rockville, MD: Substance Abuse and Mental Health Services Administration; 2015. https://www.ncbi.nlm.nih.gov/books/NBK64258/

4. Pierce M, Hayhurst K, Bird SM, et al. Insights into the link between drug use and criminality: Lifetime offending of criminally-active opiate users. *Drug and Alcohol Dependence*.

2017;179:309-316. https://www.sciencedirect.com/science/article/pii/S0376871617304106. doi: 10.1016/j.drugalcdep.2017.07.024.

5. National Drug Intelligence Center . The economic impact of illicit drug use on American society.United States Department of Justice; Washington, DC: 2011. Retrieved from http://www.justice.gov/archive/ndic/pubs44/44731/44731p.pdf.

6. Rounsaville BJ, Petry NM, Carroll KM. Single versus multiple drug focus in substance abuse clinical trials research. *Drug and Alcohol Dependence*. 2003;70(2):117-125.

7. Ogbu UC, Lotfipour S, Chakravarthy B. Polysubstance abuse: Alcohol, opioids and benzodiazepines require coordinated engagement by society, patients, and physicians. *The western journal of emergency medicine*. 2015;16(1):76-79.

http://www.ncbi.nlm.nih.gov/pubmed/25671013. doi: 10.5811/westjem.2014.11.24720.

8. Sachdeva A, Choudhary M, Chandra M. Alcohol withdrawal syndrome: Benzodiazepines and beyond. *Journal of clinical and diagnostic research : JCDR*. 2015;9(9):VE01. http://www.ncbi.nlm.nih.gov/pubmed/26500991.

9. Betts KS, McIlwraith F, Dietze P, et al. Can differences in the type, nature or amount of polysubstance use explain the increased risk of non-fatal overdose among psychologically distressed people who inject drugs? *Drug and Alcohol Dependence*. 2015;154:76-84. https://www.clinicalkey.es/playcontent/1-s2.0-S0376871615003191. doi: 10.1016/j.drugalcdep.2015.06.020.

10. Bauer LK, Brody JK, León C, Baggett TP. Characteristics of homeless adults who died of drug overdose: A retrospective record review. *Journal of Health Care for the Poor and*

Underserved. 2016;27(2):846-859. https://muse.jhu.edu/article/617500. doi: 10.1353/hpu.2016.0075.

11. Salom CL, Betts KS, Williams GM, Najman JM, Alati R. Predictors of comorbid polysubstance use and mental health disorders in young adults-a latent class analysis. Addiction. 2016;111(1):156-164. doi: 10.1111/add.13058.

12. Bhalla I, Stefanovics E, Rosenheck R. Clinical epidemiology of single versus multiple substance use disorders: Polysubstance use disorder. Medical Care. 2017;55 Suppl 9 Suppl 2:S32. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=n&CSC=Y&PAGE=fulltext&D=ovft&AN=00005650-201709001-00008. doi: 10.1097/MLR.0000000000000731.

13. Joseph J Palamar, Patricia Acosta, Charles M Cleland. Attitudes and beliefs about new psychoactive substance use among electronic dance music party attendees. *Substance Use & Misuse.* 2018;53(3):381. https://search.proquest.com/docview/2009351420. doi: 10.1080/10826084.2017.1327980.

14. Fermín Fernández-Calderón, Charles M Cleland, Joseph J Palamar. Polysubstance use profiles among electronic dance music party attendees in new york city and their relation to use of new psychoactive substances. *Addictive Behaviors.* 2018;78:85. https://search.proquest.com/docview/2019060192.

15. Walton GRT, Hayashi K, Bach P, et al. The impact of benzodiazepine use on mortality among polysubstance users in vancouver, canada. *Public Health Reports*. 2016;131(3):491-

499. http://journals.sagepub.com/doi/full/10.1177/003335491613100315. doi: 10.1177/003335491613100315.

16. Gjersing L, Bretteville-Jensen AL. Patterns of substance use and mortality risk in a cohort of 'hard-to-reach' polysubstance users. *Addiction*. 2018;113(4):729-739. https://onlinelibrary.wiley.com/doi/abs/10.1111/add.14053. doi: 10.1111/add.14053.

17. Das, Jai K., M.D., M.B.A, Salam RA, M.Sc, Arshad A, M.B.B.S, Finkelstein Y, M.D, Bhutta ZA, Ph.D. Interventions for adolescent substance abuse: An overview of systematic reviews. Journal of Adolescent Health. 2016;59(4):S75. https://www.clinicalkey.es/playcontent/1-s2.0-S1054139X16301677. doi: 10.1016/j.jadohealth.2016.06.021.

18. Tomczyk S, Isensee B, Hanewinkel R. Latent classes of polysubstance use among adolescents-a systematic review. *Drug and alcohol dependence*. 2016;160:12. https://www.clinicalkey.es/playcontent/1-s2.0-S0376871615017895. doi: 10.1016/j.drugalcdep.2015.11.035.

19. Robinson SM, Sobell LC, Sobell MB, Leo GI. Reliability of the timeline followback for cocaine, cannabis, and cigarette use. Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors. 2014;28(1):154-162. http://www.ncbi.nlm.nih.gov/pubmed/23276315. doi: 10.1037/a0030992.

20. Hjorthøj CR, Hjorthøj AR, Nordentoft M. Validity of timeline follow-back for self-reported use of cannabis and other illicit substances — systematic review and meta-analysis. Addictive Behaviors. 2011;37(3):225-233. https://www.clinicalkey.es/playcontent/1-s2.0-S030646031100387X. doi: 10.1016/j.addbeh.2011.11.025.

21. Roy M, Dum M, Sobell LC, et al. Comparison of the quick drinking screen and the alcohol timeline followback with outpatient alcohol abusers. Substance Use & Misuse. 2008;43(14):2116-2123. http://www.ingentaconnect.com/content/apl/lsum/2008/00000043/00000014/art00006. doi: 10.1080/10826080802347586.

22. 1. Kent P, Jensen RK, Kongsted A. A comparison of three clustering methods for finding subgroups in MRI, SMS or clinical data: SPSS TwoStep cluster analysis, latent gold and SNOB. BMC medical research methodology. 2014;14(1):113. http://www.ncbi.nlm.nih.gov/pubmed/25272975. doi: 10.1186/1471-2288-14-113.

23. Tao Shi, Steve Horvath. Unsupervised learning with random forest predictors. *Journal of Computational and Graphical Statistics.* 2006;15(1):118-138. http://www.jstor.org/stable/27594168. doi: 10.1198/106186006X94072.

24. Shi T, Seligson D, Belldegrun AS, Palotie A, Horvath S. Tumor classification by tissue microarray profiling: Random forest clustering applied to renal cell carcinoma. *Modern Pathology.* 2005;18(4):547-557. http://dx.doi.org/10.1038/modpathol.3800322. doi: 10.1038/modpathol.3800322.

25. NIDA CTN. Instrument: Timeline Followback Method Assessment. https://cde.drugabuse.gov/instrument/d89c8e23-16e5-625a-e040-bb89ad43465d. Accessed on April 20, 2018.

26. National Institute of Alcohol Abuse and Alcoholism. Drinking Levels Defined. https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking. Accessed on April 20, 2018.

27. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32. https://search.proquest.com/docview/757027982. doi: 1010933404324.

28. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Vol 103. New York: Springer; 2013. http://www.fachportal-paedagogik.de/fis_bildung/suche/fis_set.html?FId=1040380. 10.1007/978-1-4614-7138-7.

29. Cai Y, Sun Y. ESPRIT-tree: Hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time. *Nucleic Acids Research.* 2011;39(14):e95. http://www.ncbi.nlm.nih.gov/pubmed/21596775. doi: 10.1093/nar/gkr349.

30. Derek Y. Chiang, Patrick O. Brown, Michael B. Eisen. Visualizing associations between genome sequences and gene expression data using genome-mean expression profiles. *Bioinformatics.* 2001;17(1):S55. http://www.ncbi.nlm.nih.gov/pubmed/11472992. doi: 10.1093/bioinformatics/17.suppl_1.S49.

31. Juan Nunez-Iglesias, Ryan Kennedy, Toufiq Parag, Jianbo Shi, Dmitri B Chklovskii. Machine learning of hierarchical clustering to segment 2D and 3D images. PLoS One. 2013;8(8):e71715. http://www.ncbi.nlm.nih.gov/pubmed/23977123. doi: 10.1371/journal.pone.0071715.

32. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: The dynamic tree cut package for R. *Bioinformatics.* 2008;24(5):719-720. http://www.ncbi.nlm.nih.gov/pubmed/18024473. doi: 10.1093/bioinformatics/btm563

33. Swihart B, Caffo B, James B, Strand M, Schwartz B, Punjabi N. Lasagna plots: A saucy alternative to spaghetti plots. *Epidemiology.* 2010;21(5):621-625. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=n&CSC=Y&PAGE=fulltext&D=ovft&AN=00001648-201009000-00015. doi: 10.1097/EDE.0b013e3181e5b06a.

34. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in medicine.* 2007;26(11):2389-2430. http://www.narcis.nl/publication/RecordID/oai:pure.amc.nl:publications%2F8b5d7610-1185-4388-b749-52d8a0f52fa7. doi: 10.1002/sim.2712.

Appendices

Table1.
Demographic characteristics of the participants

Variables	N=439
Age, median (IQR)	31 (25, 44)
Race (%)	
Black	438 (99.8)
White	1 (0.2)
Hispanic or Latina (%)	
No	422 (96.1)
Yes	17 (3.9)
Education (%)	
< High school	133 (30.3)
High school/GED	131 (29.8)
>High school	175 (39.9)
Marital Status (%)	
Single (never married)	316 (72.0)
Married	24 (5.5)
Separated	26 (5.9)
Divorced	47 (10.7)
Widowed	8 (1.8)
Cohabiting	18 (4.1)
Living with children (%)	
No children	140 (31.9)
Having children but not living together	86 (19.6)
Living with children	213 (48.5)
Living Situation (%)	
Stable housing	232(52.8)
Unstable housing	207 (47.2)
Employment (%)	
Full-time employed	361 (82.2)
Part-time employed	29 (6.6)
Student	9 (2.1)
Unemployed	40 (9.1)
Income (%)	
$ 0-5,000	215 (49.0)
$ 5,001-15,000	147 (33.5)
>$15,000	77 (17.5)
Time to first cigarette after waking (%)	
Do not smoke	193 (44.0)
Within 5 minutes	119 (27.1)
6 - 30 minutes	71 (16.2)
31-60 minutes	21 (4.8)
Over 60 minutes	35 (8.0)
Cigarettes per day (%)	
0	193(44.0)
<10	132 (30.1)
≥10	114 (26.0)
HIV positive (%)	
No	401 (91.3)
Yes	22 (5.0)
Unsure	16 (3.6)

Table2.
Proportions of days with certain number of substance use for each number of substances reported for the entire 30-day period

Number of substances reported	Proportion of days with zero substances used	Proportion of days with one substance used	Proportion of day with two substances used	Proportion of days with three substances used	Proportion of days with four substances used	Proportion of days with five substances used
1	0.83 (0.63, 0.90)	0.17 (0.10, 0.37)	-	-	-	-
2	0.47 (0.07, 0.73)	0.40 (0.20, 0.63)	0.07 (0.03, 0.20)	-	-	-
3	0.37 (0, 0.63)	0.25 (0.17, 0.48)	0.15 (0.07, 0.32)	0.03 (0, 0.08)	-	-
4	0.43 (0.37, 0.60)	0.17 (0.13, 0.20)	0.2 (0.13, 0.30)	0.07 (0.03, 0.07)	0 (0, 0.03)	-
5	0.10 (0.07, 0.13)	0.28 (0.07, 0.50)	0.32 (0.30, 0.33)	0.17 (0.07, 0.27)	0.10 (0.03, 0.17)	0.03 (0, 0.07)

* Median (interquartile range)

Table3.
Substance use according to clusters on the day level

Cluster	% (n)	Drinks	Binge	Marijuana	Cocaine	Heroin	Prescription Drugs
Cluster1	49.36% (n=6461)	0 (0, 0)	0	0	0	0	0
Cluster2	13.14% (n=1720)	0 (0, 0)	0	1	0	0	0
Cluster3	10.18% (n=1332)	5.3 (4.4, 6.8)	1	0	0	0	0
Cluster4	6.05% (n=792)	6.6 (4.7, 9.3)	1	1	0	0	0
Cluster5	4.19% (n=548)	2.2 (1.8, 2.6)	0	0	0	0	0
Cluster6	3.73% (n=488)	19 (16.1, 25.2)	1	0.39	0	0	0
Cluster7	3.54% (n=463)	9.4 (6.7, 16)	0.98	0.24	0.61	0.37	0.17
Cluster8	3.51% (n=460)	10.7 (9.8, 12)	1	0	0	0	0
Cluster9	2.48% (n=325)	2 (1.1, 2.4)	0	1	0	0	0
Cluster10	1.12% (n=147)	1.2 (1, 1.3)	0	0	0	0	0
Cluster11	1.00% (n=131)	0 (0, 0)	0	0	0.60	0.40	0
Cluster12	0.86% (n=113)	0 (0, 1.6)	0	0.38	0.42	0.30	0.38
Cluster13	0.84% (n=110)	0.8 (0.7, 0.9)	0	0	0	0	0

* Values of drinks refer to the median (interquartile range) number of standard drinks. Values of other variables refer the proportion of taking the substance.

Table4.
Transition probability matrix of polysubstance use clusters on the day level

	Abs	Mi to mo D	B	M	M+mi to mo D	B+sM	B+M	Sc+sH	Mi to mo D+sM+sC+ sH+sRx	B+sM+sC +sH+sRx
Abs	0.81	0.04	0.08	0.03	0.01	0.01	0.01	0.00	0.00	0.01
Mi to mo D	0.34	0.49	0.10	0.02	0.01	0.01	0.01	0.00	0.00	0.00
B	0.30	0.04	0.58	0.02	0.00	0.02	0.02	0.00	0.00	0.01
M	0.12	0.01	0.02	0.70	0.05	0.02	0.07	0.00	0.00	0.00
M+mi to mo D	0.13	0.03	0.03	0.22	0.48	0.02	0.09	0.00	0.00	0.00
B+sM	0.19	0.03	0.08	0.05	0.02	0.58	0.04	0.00	0.00	0.01
B+M	0.13	0.01	0.05	0.15	0.04	0.03	0.58	0.00	0.00	0.01
Sc+sH	0.27	0.02	0.05	0.01	0.00	0.00	0.01	0.54	0.02	0.08
Mi to mo D+sM+sC+sH+sRx	0.20	0.03	0.07	0.06	0.01	0.04	0.02	0.01	0.52	0.04
B+sM+sC+sH+sRx	0.16	0.02	0.04	0.02	0.00	0.02	0.01	0.01	0.03	0.69

* Abs, abstinence; D, drinking; B, binge; M, marijuana; C, cocaine; H, heroin; Rx, prescription drugs; s, some; mi, mild; mo, moderate.
The groups corresponded to the original clusters as follows. Abs: Cluster1; Mi to mo D: Cluster5, Cluster10, and Cluster13. B: Cluster3, and Cluster8; M:Cluster2; M+mi to mo D:Cluster9; B+sM:Cluster6; B+M: Cluster4; SC+sH: Cluster11; Mi to mo D+sM+sC+sH+sRx: Cluster12; B+sM+sC+sH+sRx: Cluster7.
The name of the row referred to pattern on the first day and the name of the column referred to the pattern on the second day.

Table 5.
Mean sojourn times for each cluster at the day level

	Estimates	SE	95% Confidence Interval	
Abs	3.89	0.10	3.69	4.10
Mi to mo D	1.37	0.06	1.25	1.50
B	1.73	0.06	1.62	1.85
M	2.57	0.13	2.32	2.85
M+mi to mo D	1.32	0.10	1.14	1.54
B+sM	1.82	0.14	1.56	2.12
B+M	1.74	0.11	1.54	1.97
Sc+sH	1.64	0.93	0.54	4.98
Mi to mo D+sM+sC+sH+sRx	1.52	0.28	1.07	2.18
B+sM+sC+sH+sRx	2.69	0.39	2.02	3.57

* SE, standard error.
Abs, abstinence; D, drinking; B, binge; M, marijuana; C, cocaine; H, heroin; Rx, prescription drugs; s, some; mi, mild; mo, moderate.

Table6.
Probability matrix of transitioning to each other clusters on the day level

	Abs	Mi to mo D	B	M	M+mi to mo D	B+sM	B+M	Sc+sH	Mi to mo D+sM+sC+ sH+sRx	B+sM+sC +sH+sRx
Abs	-	0.24	0.43	0.12	0.03	0.05	0.06	0.01	0.02	0.04
Mi to mo D	0.69	-	0.21	0.03	0.02	0.02	0.01	0.00	0.01	0.00
B	0.72	0.12	-	0.03	0.00	0.06	0.05	0.00	0.00	0.02
M	0.33	0.03	0.05	-	0.23	0.07	0.26	0.00	0.02	0.01
M+mi to mo D	0.20	0.06	0.03	0.49	-	0.02	0.19	0.00	0.01	0.01
B+sM	0.43	0.07	0.20	0.11	0.04	-	0.12	0.00	0.01	0.01
B+M	0.26	0.01	0.11	0.39	0.12	0.07	-	0.00	0.00	0.03
Sc+sH	0.60	0.03	0.08	0.00	0.01	0.00	0.01	-	0.06	0.21
Mi to mo D+sM+sC+sH+sRx	0.37	0.08	0.15	0.14	0.00	0.10	0.03	0.04	-	0.09
B+sM+sC+sH+sRx	0.49	0.08	0.12	0.07	0.00	0.06	0.02	0.03	0.13	-

* Abs, abstinence; D, drinking; B, binge; M, marijuana; C, cocaine; H, heroin; Rx, prescription drugs; s, some; mi, mild; mo, moderate.
The name of the row referred to pattern on the first day and the name of the column referred to the pattern on the second day.

Table7.
Parameter estimates (hazard ratios) for multistate model

	First cigarette ≤60 min after waking up	First cigarette >60 min after waking up	Age
Abs - Abs			
Abs - Mi to mo D	0.63 (0.50,0.78)	0.69(0.48,1.00)	0.99 (0.98,0.99)
Abs - B	1.49 (1.26,1.75)	1.16 (0.88,1.53)	1.00 (1.00,1.01)
Abs - M	1.95 (1.47,2.57)	0.64 (0.33,1.26)	0.94 (0.92,0.95)
Abs - M+mi to mo D	3.14 (1.69,5.83)	1.48 (0.46,4.78)	0.93 (0.90,0.96)
Abs - B+sM	6.02 (3.67,9.89)	3.25 (1.52,6.95)	0.95 (0.93,0.97)
Abs - B+M	3.43 (2.15,5.46)	1.23 (0.49,3.11)	0.97 (0.95,0.99)
Abs - SC+sH	15.60 (3.96,61.52)	4.09 (0.65,25.79)	1.05 (1.02,1.08)
Abs – Mi to mo D+sM+sC+sH+sRx	4.40 (2.04,9.47)	0.49 (0.05,4.68)	1.00 (0.98,1.03)
Abs - B+sM+sC+sH+sRx	7.06 (3.79,13.13)	1.57 (0.53,4.64)	1.04 (1.03,1.06)
Mi to mo D - Abs	0.92 (0.73,1.15)	0.89 (0.61,1.30)	1.00 (0.99,1.01)
Mi to mo D - Mi to mo D			
Mi to mo D - B	1.03 (0.68,1.54)	0.97 (0.48,1.95)	1.00 (0.99,1.02)
Mi to mo D - M	4.45 (1.98,9.98)	2.79 (0.83,9.42)	0.91 (0.87,0.96)
Mi to mo D - M+mi to mo D	1.35 (0.36,5.06)	0.23 (0.00,33.63)	1.02 (0.97,1.08)
Mi to mo D - B+sM	1.42 (0.48,4.17)	0.20 (0.00,10.38)	0.94 (0.89,1.00)
Mi to mo D - B+M	4.74 (1.01,22.11)	2.86 (0.28,29.30)	0.93 (0.86,1.01)
Mi to mo D - SC+sH	22.78 (0.37,1392.00)	0.76 (0.00,$4.85*10^5$)	1.10 (1.02,1.20)
Mi to mo D – Mi to mo D+sM+sC+sH+sRx	3.30 (0.49,22.05)	0.51 (0.00,256.60)	1.05 (0.97,1.12)
Mi to mo D - B+sM+sC+sH+sRx	8.70 (0.60,125.50)	0.60 (0.00,5069.00)	1.13 (1.04,1.23)
B - Abs	0.90 (0.77,1.05)	1.06 (0.79,1.40)	0.99 (0.98,1.00)
B - Mi to mo D	0.52 (0.35,0.77)	0.81 (0.40,1.63)	0.98 (0.96,1.00)
B - B			
B - M	1.17 (0.53,2.59)	0.53 (0.06,4.70)	0.95 (0.92,0.99)
B - M+mi to mo D	26.20 (0.09,7702.00)	0.65 (0.00,$7.91*10^6$)	1.01 (0.92,1.09)
B - B+sM	2.15 (1.21,3.83)	2.00 (0.78,5.13)	0.99 (0.97,1.01)
B - B+M	2.49 (1.35,4.57)	1.21 (0.34,4.35)	0.97 (0.95,0.99)
B - SC+sH	22.76 (0.61,851.30)	0.56 (0.00,2391.00)	1.16 (1.06,1.26)
B – Mi to mo D+sM+sC+sH+sRx	0.37 (0.058,2.41)	0.31 (0.00,21.25)	1.03 (0.96,1.11)
B - B+sM+sC+sH+sRx	2.25 (0.92,5.50)	0.92 (0.16,5.40)	1.05 (1.02,1.09)
M - Abs	1.10 (0.83,1.47)	0.44 (0.20,1.00)	1.02 (1.00,1.03)
M - Mi to mo D	0.82 (0.35,1.93)	3.07 (1.05,8.99)	0.98 (0.93,1.03)
M - B	1.41 (0.64,3.09)	0.56 (0.07,4.32)	1.03 (0.99,1.08)

M - M			
M - M+mi to mo D	0.89 (0.63,1.27)	0.71 (0.33,1.53)	1.02 (1.00,1.04)
M - B+sM	1.89 (1.07,3.35)	0.78 (0.18,3.37)	0.98 (0.94,1.02)
M - B+M	1.28 (0.96,1.72)	0.50 (0.21,1.16)	0.98 (0.96,1.00)
M - SC+sH	-	-	-
M – Mi to mo D+sM+sC+sH+sRx	3.54 (0.73,17.09)	0.26 (0.00,72.60)	1.07 (1.00,1.14)
M - B+sM+sC+sH+sRx	5.99 (0.68,52.98)	0.44 (0.00,1363.00)	1.00 (0.90,1.12)
M+mi to mo D - Abs	0.53 (0.29,0.98)	0.72 (0.21,2.48)	1.01 (0.98,1.04)
M+mi to mo D - Mi to mo D	0.98 (0.32,2.97)	1.01 (0.10,10.63)	1.02 (0.96,1.07)
M+mi to mo D - B	2.44 (0.50,11.77)	0.46 (0.00,278.30)	1.05 (0.99,1.11)
M+mi to mo D - M	0.95 (0.67,1.36)	0.94 (0.47,1.87)	0.98 (0.96,1.00)
M+mi to mo D - M+mi to mo D			
M+mi to mo D - B+sM	0.82 (0.15,4.49)	0.42 (0.00,104.20)	1.04 (0.96,1.12)
M+mi to mo D - B+M	1.31 (0.74,2.30)	0.18 (0.00,2.33)	0.99 (0.96,1.02)
M+mi to mo D - SC+sH	-	-	-
M+mi to mo D – Mi to mo D+sM+sC+sH+sRx	0.82 (0.05,14.07)	0.86 (0.00, 4109.00)	1.08 (0.97,1.20)
M+mi to mo D - B+sM+sC+sH+sRx	0.92 (0.03,25.10)	0.73 (0.00,3574.00)	1.03 (0.89,1.20)
B+sM - Abs	0.82 (0.52,1.30)	2.12(1.09,4.15)	0.98 (0.97,1.00)
B+sM - Mi to mo D	0.15 (0.05,0.41)	0.03 (0.00,5.60)	1.06 (1.01,1.11)
B+sM - B	1.29 (0.59,2.81)	1.33 (0.38,4.62)	1.03 (1.01,1.06)
B+sM - M	0.44 (0.25,0.80)	0.18 (0.01,2.33)	0.91 (0.88,0.96)
B+sM - M+mi to mo D	0.24 (0.06,0.97)	1.39 (0.22,8.68)	1.00 (0.94,1.07)
B+sM - B+sM			
B+sM - B+M	1.29 (0.57,2.89)	1.54 (0.35,6.72)	0.96 (0.93,0.99)
B+sM - SC+sH	-	-	-
B+sM – Mi to mo D+sM+sC+sH+sRx	6.24 (0.04,889.50)	0.48 (0.00,1.77*10^4)	1.07 (0.99,1.15)
B+sM - B+sM+sC+sH+sRx	24.80 (0.08,7695.00)	0.36 (0.00,2.83*10^5)	1.07 (1.03,1.12)
B+M - Abs	0.71 (0.46,1.08)	1.58 (0.72,3.44)	1.04 (1.02,1.06)
B+M - Mi to mo D	0.39 (0.07,2.08)	0.57 (0.00,68.24)	1.01 (0.93,1.10)
B+M - B	1.66 (0.83,3.30)	0.79 (0.12,5.22)	1.04 (1.01,1.07)
B+M - M	1.02 (0.76,1.37)	1.03 (0.41,2.59)	0.97 (0.96,0.99)
B+M - M+mi to mo D	0.64 (0.38,1.07)	0.04 (0.00,19.42)	1.01 (0.98,1.03)
B+M - B+sM	1.26 (0.61,2.61)	1.24 (0.16,9.67)	0.99 (0.95,1.02)
B+M - B+M			
B+M - SC+sH	-	-	-
B+M – Mi to mo D+sM+sC+sH+sRx	2.27(0.02,319.70）	0.95 (0.00,2.23*10^7)	0.98 (0.77,1.25)
B+M - B+sM+sC+sH+sRx	1.18 (0.37,3.82)	0.20 (0.00,20.87)	1.08 (1.03,1.13)
SC+sH - Abs	0.42 (0.04,4.55)	0.44 (0.03,6.31)	1.02 (0.97,1.07)
SC+sH - Mi to mo D	1.62 (0.00,6.29*10^6)	0.57 (0.00,6.62*10^6)	1.02 (0.87,1.20)

SC+sH - B	1.95 (0.00,8.41*10^{5})	0.67 (0.00,4.08*10^{6})	0.93 (0.80,1.07)
SC+sH - M	-	-	-
SC+sH - M+mi to mo D	1.17 (0.00,7.06*10^{8})	0.78 (0.00,9.22*10^{8})	1.08 (0.85,1.36)
SC+sH - B+sM	-	-	-
SC+sH - B+M	1.13 (0.00,3.35*10^{11})	0.76 (0.00,1.48*10^{12})	1.01 (0.72,1.41)
SC+sH - SC+sH			
SC+sH – Mi to mo D+sM+sC+sH+sRx	3.34 (0.00,4.30*10^{5})	0.24 (0.00,1.30*10^{5})	1.03 (0.94,1.12)
SC+sH - B+sM+sC+sH+sRx	0.70 (0.00,394.90)	2.32 (0.00,1936.00)	0.94 (0.86,1.04)
Mi to mo D+sM+sC+sH+sR - Abs	0.84 (0.31,2.24)	3.95 (0.47,33.37)	1.02 (0.98,1.06)
Mi to mo D+sM+sC+sH+sR - Mi to mo D	1.38 (0.16,12.18)	0.87 (0.00,1.62*10^{4})	1.02 (0.95,1.11)
Mi to mo D+sM+sC+sH+sR - B	1.27 (0.26,6.26)	0.77 (0.00,1536.00)	0.99 (0.94,1.05)
Mi to mo D+sM+sC+sH+sR - M	0.53 (0.14,2.04)	0.82 (0.00,1379.00)	0.90 (0.83,0.98)
Mi to mo D+sM+sC+sH+sR - M+mi to mo D	1.34 (0.00,1068.00)	1.09 (0.00,1.85*10^{22})	0.85 (0.53,1.36)
Mi to mo D+sM+sC+sH+sR - B+sM	0.30 (0.06,1.56)	0.80 (0.00,393.50)	1.01 (0.93,1.09)
Mi to mo D+sM+sC+sH+sR - B+M	0.60 (0.05,7.15)	1.01 (0.00,8.86*10^{7})	0.84 (0.67,1.06)
Mi to mo D+sM+sC+sH+sR - SC+sH	10.86 (0.11,1107.00)	0.96 (0.00,2.23*10^{9})	1.04 (0.97,1.11)
Mi to mo D+sM+sC+sH+sR – Mi to mo D+sM+sC+sH+sRx			
Mi to mo D+sM+sC+sH+sR - B+sM+sC+sH+sRx	4.25 (0.39,45.85)	0.86 (0.00,1.08*10^{5})	1.05 (0.99,1.11)
B+sM+sC+sH+sR - Abs	0.80 (0.36,1.77)	0.28 (0.06,1.24)	1.01 (0.99,1.04)
B+sM+sC+sH+sR - Mi to mo D	0.33 (0.06,1.65）	0.15 (0.00,5.77)	1.01 (0.95,1.08)
B+sM+sC+sH+sR - B	0.28 (0.11,0.75)	0.22 (0.03,1.44)	1.08 (1.03,1.13)
B+sM+sC+sH+sR - M	1.28 (0.27,5.99)	0.06 (0.00,4.95)	0.84 (0.78,0.91)
B+sM+sC+sH+sR - M+mi to mo D	-	-	-
B+sM+sC+sH+sR - B+sM	1.82 (0.12,27.30)	0.20 (0.00,80.72)	1.02 (0.95,1.09)
B+sM+sC+sH+sR - B+M	32.96 (0.03,4.02*10^{4})	0.14 (0.00,4.53*10^{6})	0.92 (0.86,0.97)
B+sM+sC+sH+sR - SC+sH	1.91 (0.08,46.53)	2.85 (0.08,98.62)	1.08 (1.01,1.16)
B+sM+sC+sH+sR – Mi to mo D+sM+sC+sH+sRx	0.61(0.05,0.48)	0.04 (0.00,2.51)	1.02 (0.97,1.08)
B+sM+sC+sH+sR - B+sM+sC+sH+sRx			

* Abs, abstinence; D, drinking; B, binge; M, marijuana; C, cocaine; H, heroin; Rx, prescription drugs; s, some; mi, mild; mo, moderate.
The name of the row referred to pattern on the first day and the name of the column referred to the pattern on the second day.

Table 8.
Probability matrix of transitioning to each clusters on the week level

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster 5	Cluster 6	Cluster 7	Cluster8
Cluster1	-	0.09	0.21	0.25	0.06	0.10	0.03	0.05
Cluster2	0.22	-	0.00	0.17	0.07	0.08	0.19	0.07
Cluster3	0.69	0.00	-	0.00	0.03	0.00	0.00	0.00
Cluster4	0.41	0.11	0.02	-	0.09	0.07	0.02	0.04
Cluster5	0.25	0.15	0.00	0.20	-	0.03	0.15	0.08
Cluster6	0.44	0.09	0.00	0.09	0.00	-	0.06	0.09
Cluster7	0.07	0.24	0.00	0.10	0.14	0.07	-	0.07
Cluster8	0.19	0.04	0.00	0.00	0.04	0.19	0.15	-
Cluster9	0.40	0.24	0.00	0.04	0.04	0.16	0.00	0.08
Cluster10	0.17	0.00	0.50	0.00	0.08	0.00	0.00	0.00
Cluster11	0.11	0.28	0.00	0.06	0.22	0.00	0.17	0.11
Cluster12	0.25	0.31	0.00	0.06	0.13	0.00	0.13	0.06
Cluster13	0.18	0.27	0.00	0.09	0.09	0.00	0.00	0.09
Cluster14	0.29	0.07	0.00	0.21	0.00	0.07	0.00	0.14
Cluster15	0.43	0.00	0.36	0.00	0.00	0.00	0.00	0.00

	Cluster9	Cluster10	Cluster11	Cluster12	Cluster13	Cluster14	Cluster15
Cluster1	0.11	0.04	0.00	0.02	0.02	0.03	0.02
Cluster2	0.02	0.00	0.05	0.05	0.02	0.07	0.00
Cluster3	0.00	0.17	0.00	0.00	0.00	0.00	0.11
Cluster4	0.04	0.00	0.02	0.11	0.06	0.02	0.00
Cluster5	0.03	0.00	0.03	0.08	0.00	0.03	0.00
Cluster6	0.16	0.00	0.00	0.03	0.00	0.03	0.00
Cluster7	0.03	0.00	0.17	0.07	0.00	0.03	0.00
Cluster8	0.19	0.00	0.15	0.00	0.04	0.04	0.00
Cluster9	-	0.00	0.00	0.00	0.00	0.04	0.00
Cluster10	0.00	-	0.00	0.00	0.00	0.00	0.25
Cluster11	0.00	0.00	-	0.06	0.00	0.00	0.00
Cluster12	0.00	0.00	0.00	-	0.06	0.00	0.00
Cluster13	0.09	0.00	0.00	0.00	-	0.18	0.00
Cluster14	0.00	0.00	0.07	0.00	0.14	-	0.00
Cluster15	0.00	0.21	0.00	0.00	0.00	0.00	-

* The name of the row referred to pattern on the first day and the name of the column referred to the pattern on the second day.

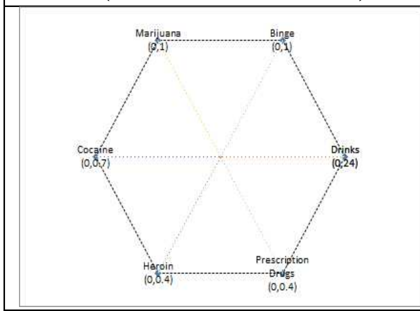
Table9.
Mean sojourn times for each cluster on the week level

	Estimates	SE	95% Confidence Interval	
Cluster 1	4.33	0.41	3.61	5.21
Cluster 2	1.56	0.20	1.21	2.01
Cluster 3	2.20	0.37	1.58	3.06
Cluster 4	1.30	0.18	0.99	1.69
Cluster 5	1.13	0.18	0.83	1.53
Cluster 6	1.25	0.22	0.88	1.77
Cluster 7	1.28	0.24	0.89	1.84
Cluster 8	1.30	0.25	0.89	1.89
Cluster 9	1.32	0.26	0.89	1.95
Cluster 10	1.42	0.41	0.80	2.49
Cluster 11	1.11	0.26	0.70	1.76
Cluster 12	1.06	0.27	0.65	1.73
Cluster 13	1.55	0.47	0.86	2.79
Custer 14	1.21	0.32	0.72	2.05
Cluster 15	1.43	0.38	0.85	2.41

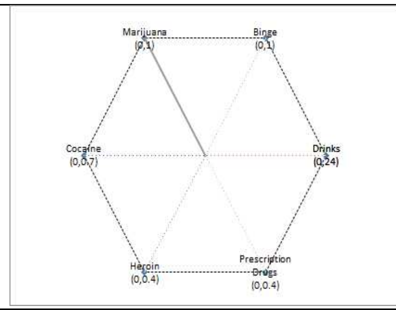
* SE, standard error.

Figure2.
Spider plot of clusters on the day level

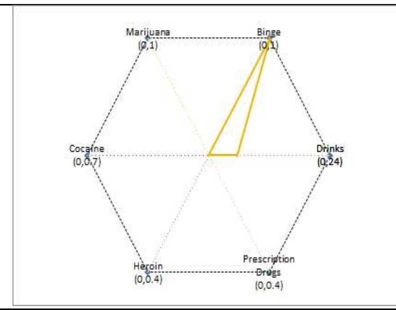
1 (n=6461, 49.36%)



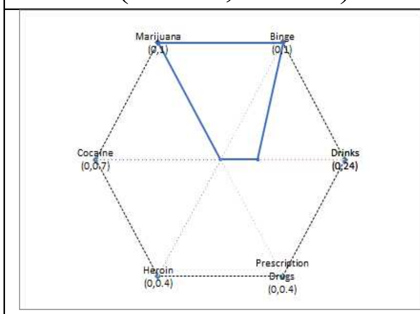
2 (n=1720, 13.14%)



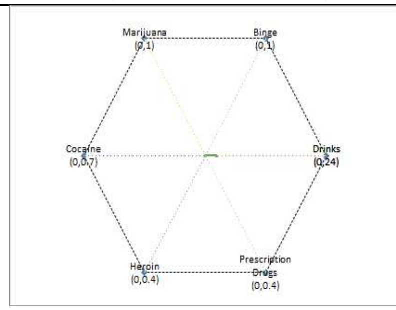
3 (n=1332, 10.18%)



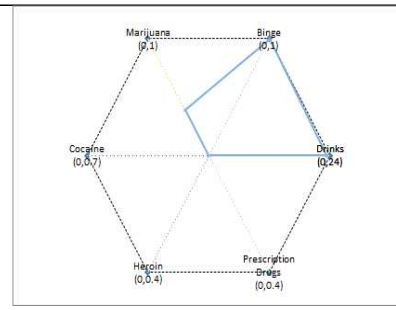
4 (n=792, 6.05%)



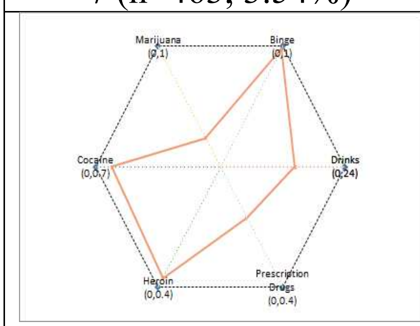
5 (n=548, 4.19%)



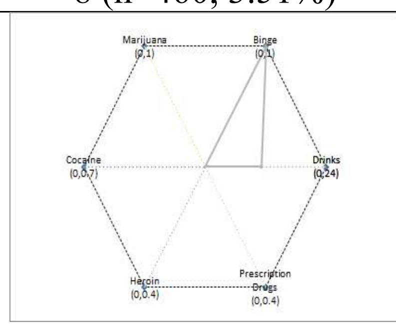
6 (n=488, 3.73%)



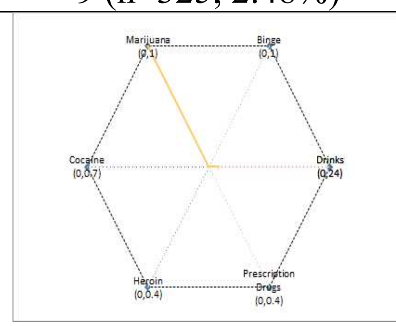
7 (n=463, 3.54%)



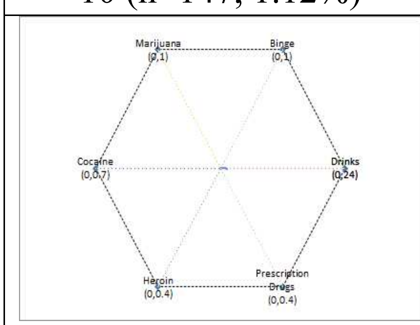
8 (n=460, 3.51%)



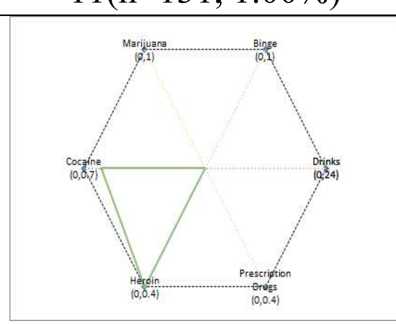
9 (n=325, 2.48%)



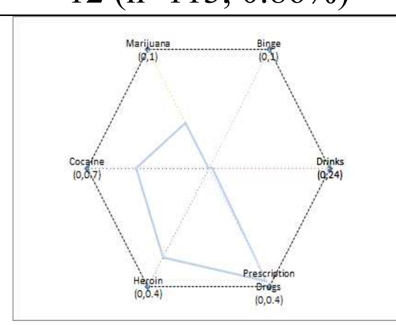
10 (n=147, 1.12%)



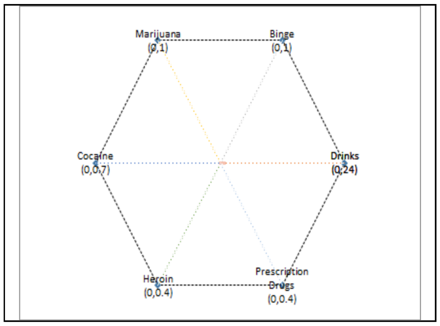
11(n=131, 1.00%)



12 (n=113, 0.86%)

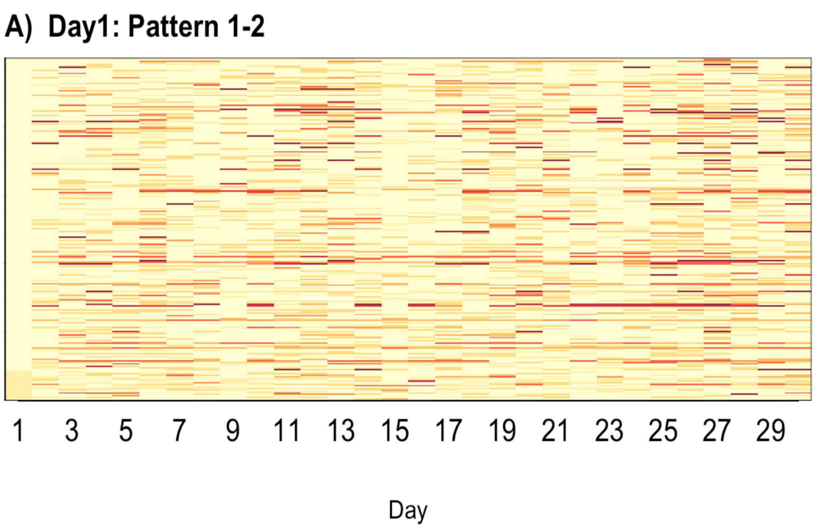


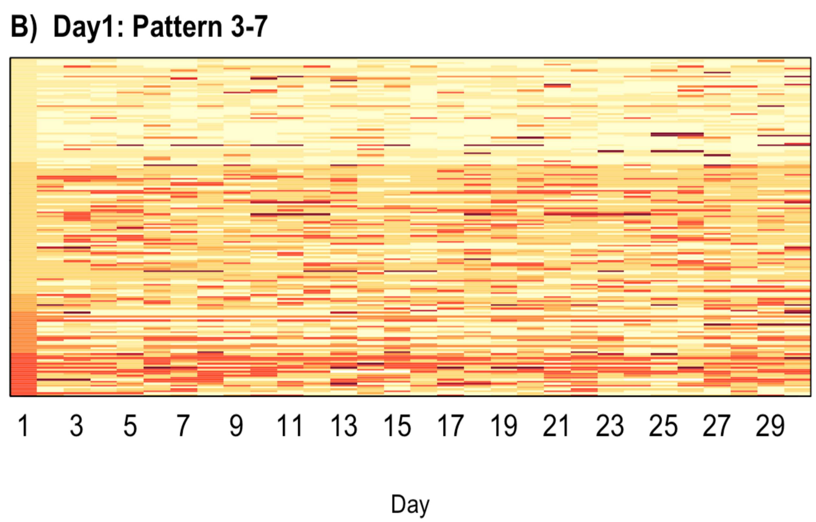
13 (n=110, 0.84%)

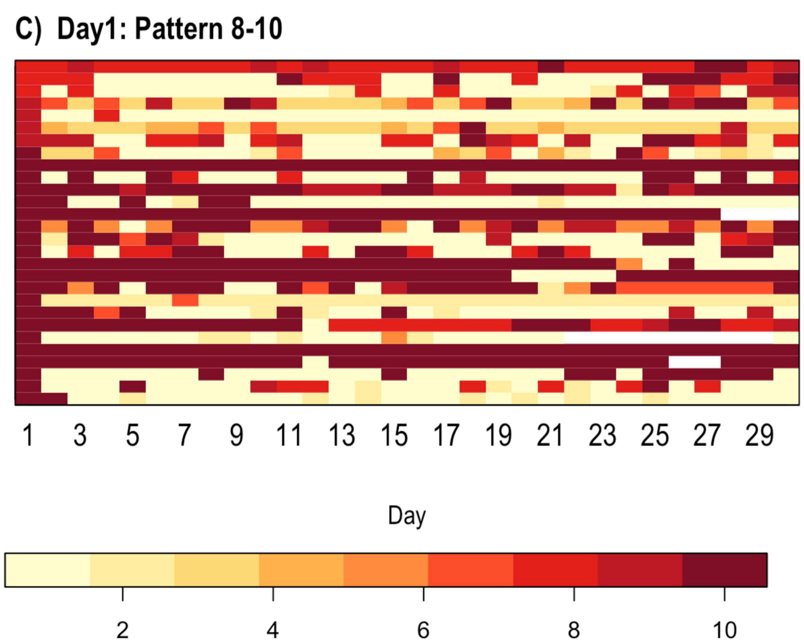


* Drinks scaled from 0 to 24, binge scaled from 0 to 1, marijuana scaled from 0 to 1, cocaine scaled from 0 to 0.7, heroin scaled from 0 to 0.4, prescription drugs scaled from 0 to 0.4.

Figure3.
Lasagna plot of daily clusters during the 30 days







* Pattern1: abstinence (n=217, 49.43%)
Pattern2: mild to moderate drinking (n=21, 4.78%)
Pattern3: binge (n=54, 12.30%)
Pattern4: marijuana (n=67, 15.26%)
Pattern5: marijuana and mild to moderate drinking (n=9, 2.05%)
Pattern6: binge and some marijuana (n=21, 4.78%)
Pattern7: binge and marijuana (n=22, 5.01%)
Pattern8: some use of cocaine and heroin (n=3, 0.68%)
Pattern9: mild to moderate drinking and some use of marijuana, cocaine, heroin, and prescription drugs (n=4, 0.91%)
Pattern10: binge and some use of marijuana, cocaine, heroin, and prescription drugs (n=21, 4.78%)

Figure4.
The 30-day polysubstance patterns of individuals that had binge drinking and some use of marijuana, cocaine, heroin, and prescription drugs for at least one day

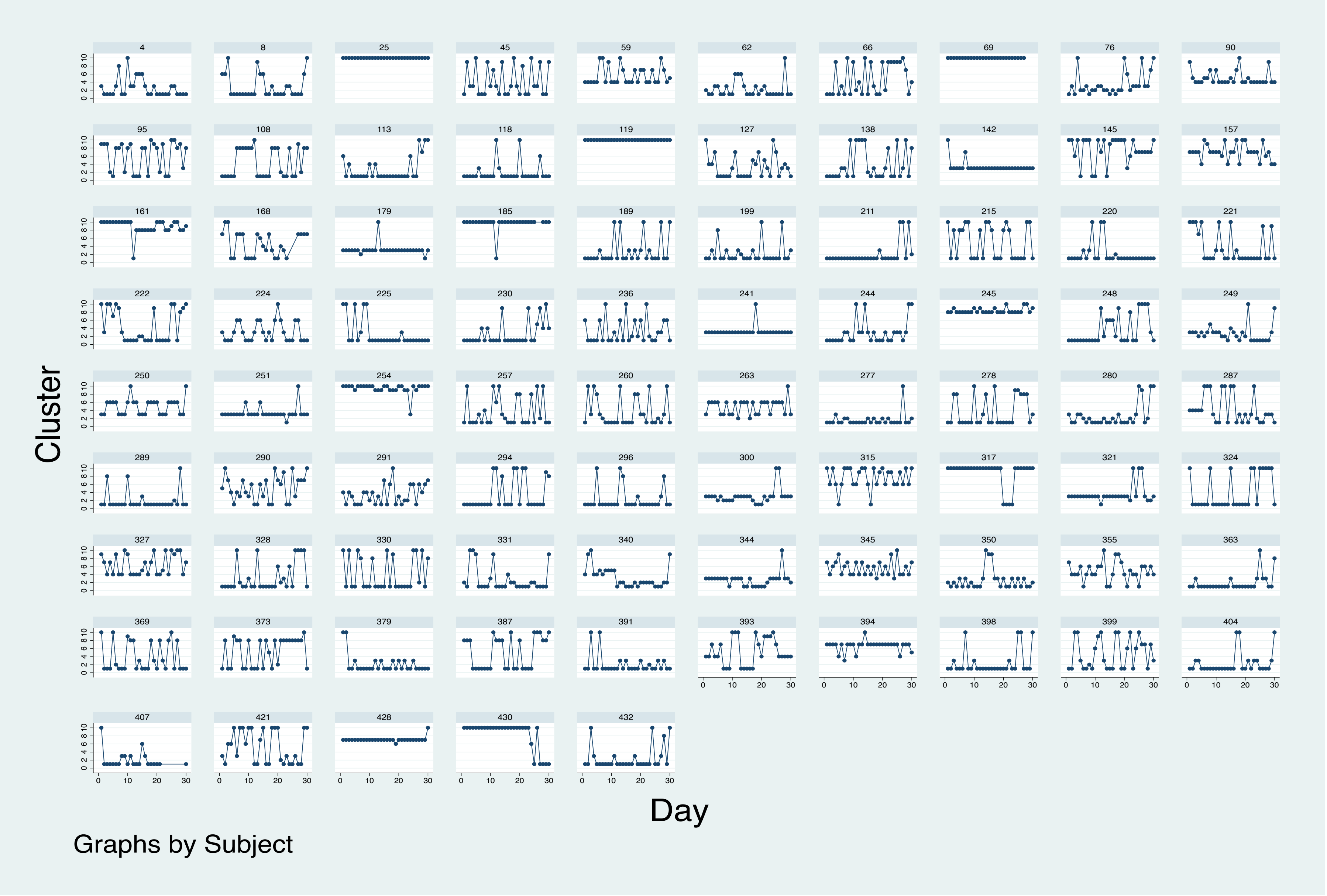
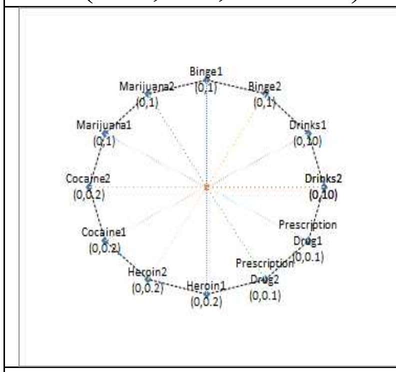
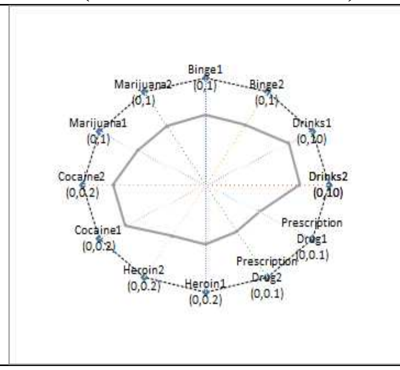


Figure5.
Spider plot of clusters for the two-day moving time window

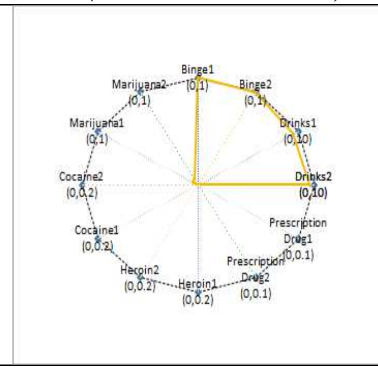
1 (n=4,493, 35.55%)



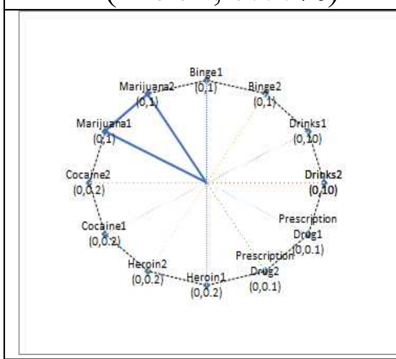
2 (n=2,220, 17.57%)



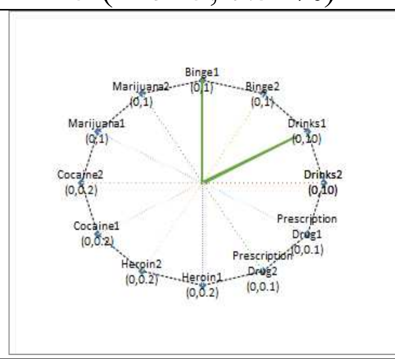
3 (n=1,060, 8.39%)



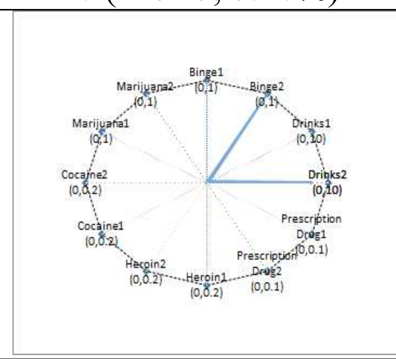
4 (n=971, 7.68%)



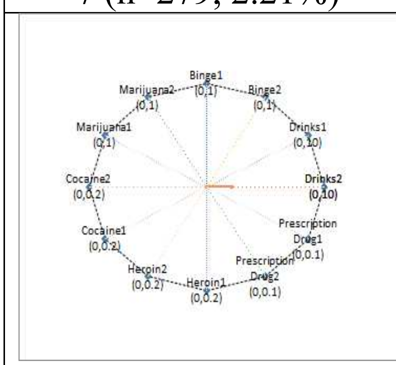
5 (n=925, 7.32%)



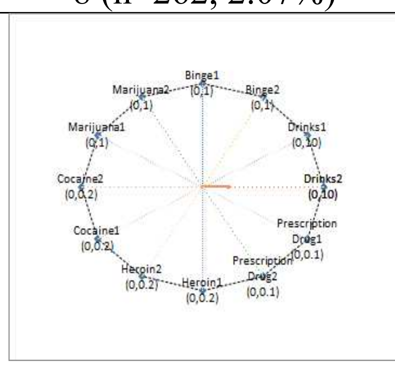
6 (n=917, 7.26%)



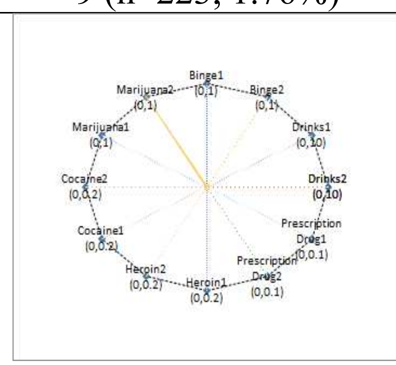
7 (n=279, 2.21%)



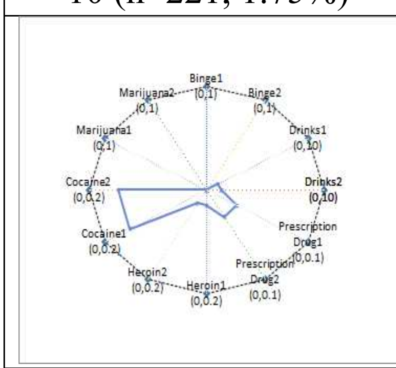
8 (n=262, 2.07%)



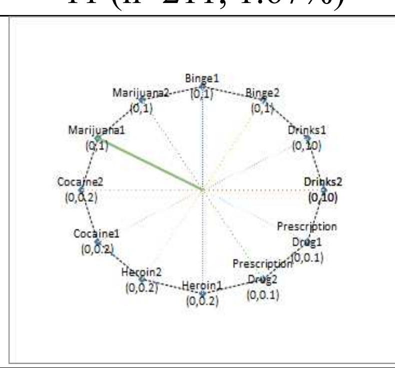
9 (n=223, 1.76%)



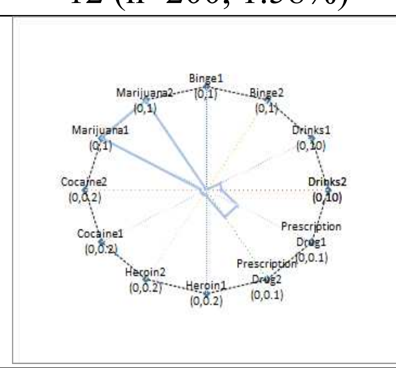
10 (n=221, 1.75%)



11 (n=211, 1.67%)



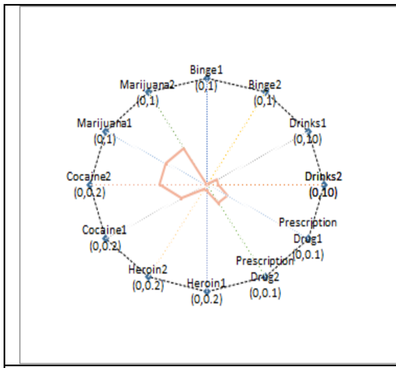
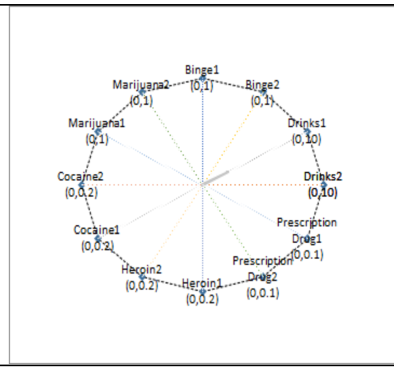
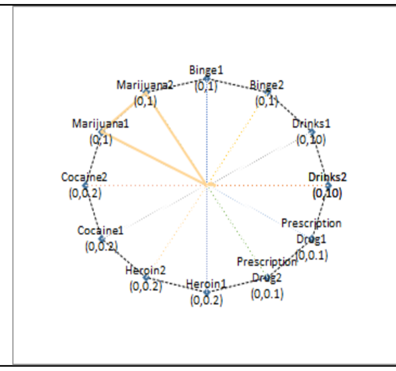
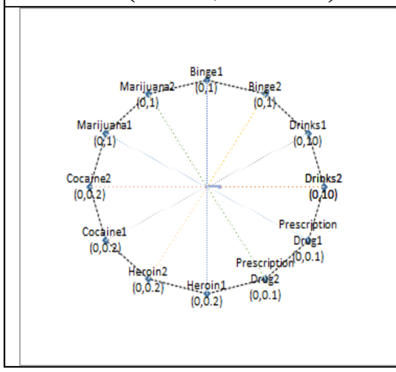
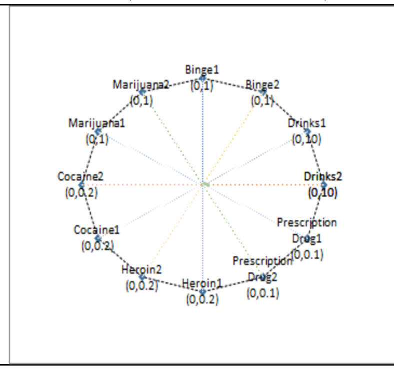
12 (n=200, 1.58%)



13 (n=193, 1.53%)

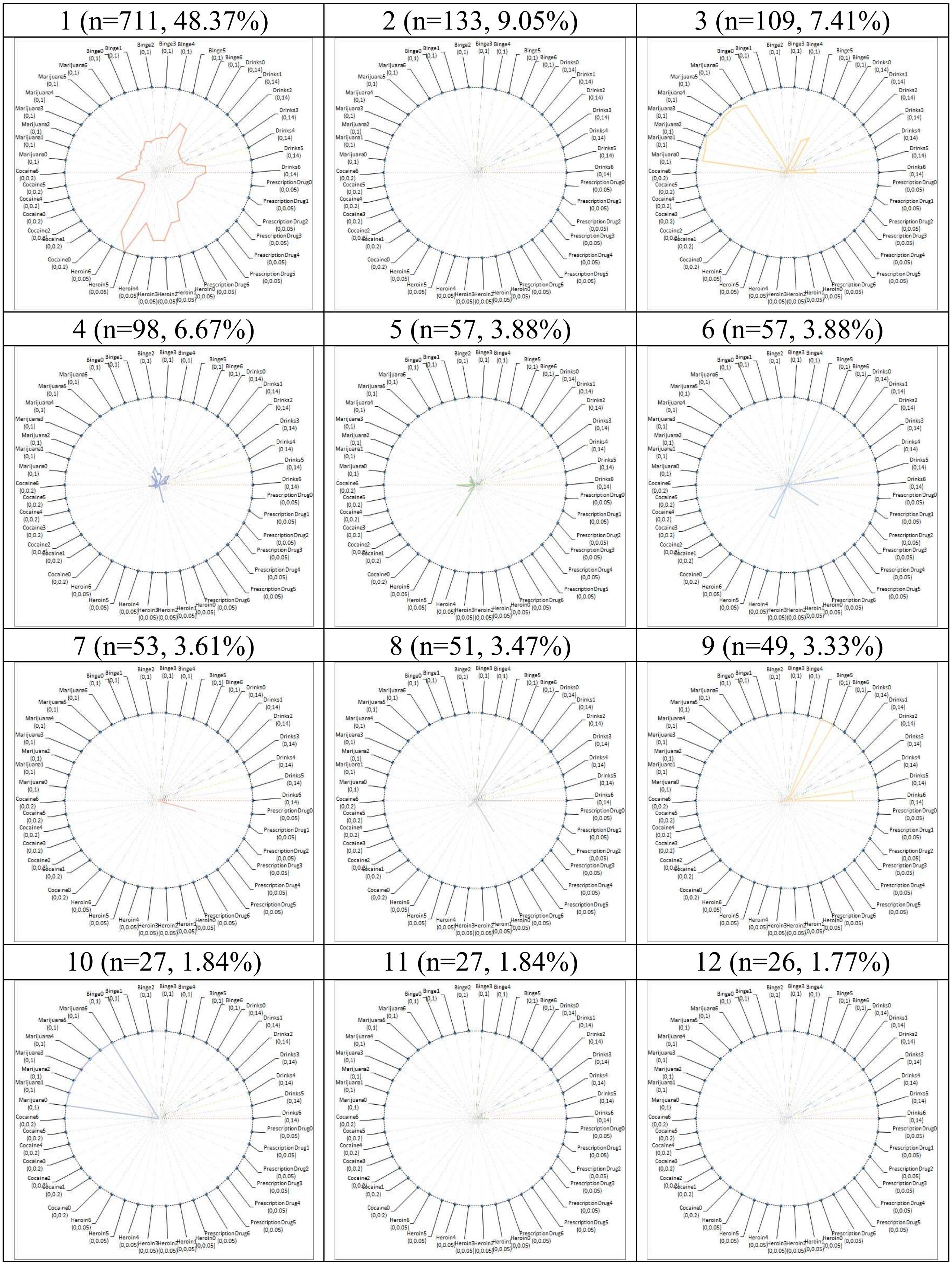
14 (n=152, 1.20%)

15 (n=133, 1.05%)

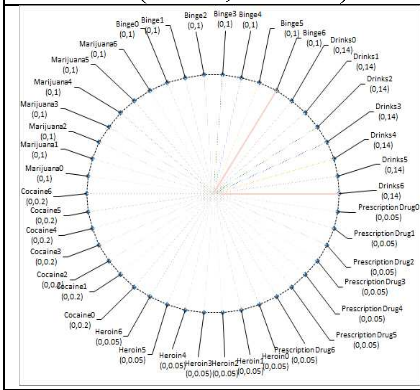
<table>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>16 (n=93, 0.74%)</td>
<td>17 (n=84, 0.66%)</td>
<td></td>
</tr>
<tr>
<td>
</td>
<td>
</td>
<td></td>
</tr>
</table>

* Substance1, the first day; substance2, the second day.
Drinks scaled from 0 to 10, binge scaled from 0 to 1, marijuana scaled from 0 to 1, cocaine scaled from 0 to 0.2, heroin scaled from 0 to 0.2, prescription drugs scaled from 0 to 0.1.

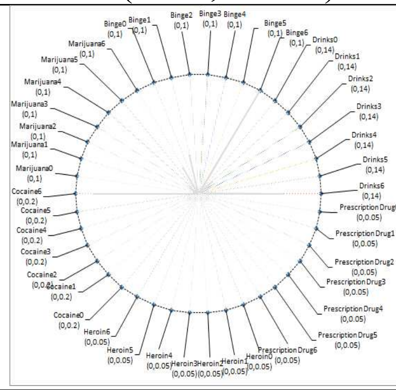
Figure6.
Spider plot of clusters on the week level



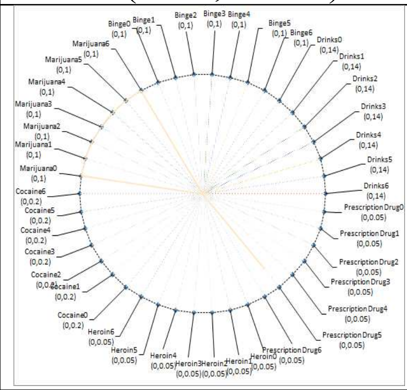
13 (n=25, 1.70%)



14 (n=24, 1.63%)



15 (n=23, 1.56%)



* Substance0, Sunday; Substance1, Monday; Substance2, Tuesday; Substance3, Wednesday; Substance4, Thursday; Substance5; Friday; Substance6, Saturday.
Drinks scaled from 0 to 14, binge scaled from 0 to 1, marijuana scaled from 0 to 1, cocaine scaled from 0 to 0.2, heroin scaled from 0 to 0.05, prescription drugs scaled from 0 to 0.05.

Curriculum Vitae

Jian Shi

Address: Unit 808, 929 N. Wolfe Street, Baltimore, MD 21205.
E-mail: jshi25@jhu.edu
Tel: 6673332011

EDUCATION

Johns Hopkins University, MD, U.S. — Sep.2016- present
Master of Science (expected in May.2018)
Major: Epidemiology (General Epidemiology and Methodology Track)

Beijing University of Chinese Medicine, Beijing, China — Sep.2011-Jul.2016
Bachelor of Arts
Major: English (Medical Science)

INTERNSHIP EXPERIENCE

International Union against Tuberculosis and Lung Disease (the Union), Beijing, China — Aug.2015-Nov.2015
Full-time Intern (Researcher)

- Project involved: Tobacco Control
- Took charge of the literature research and wrote a literature review
- Engaged in the preparation of training projects
- Attended MPOWER/FCTC Workshop in China
- Searched for tobacco control publicity videos, appropriate to be played during the trainings, and contacted the related organizations for usage authorization
- Joined tobacco control meetings, such as China ICC meeting at WHO China Office

Beijing Chaoyang Health and Family Planning Commission, Beijing, China — Jul.2015-Aug.2015
Full-time Intern, Disease Control and Prevention Division

- Did literature research in the studies of serious epidemic diseases, and completed a literature review of Main Problems of Tuberculosis Prevention and Treatment in China by referring to CNKI database
- Archived government documents
- Attended Chinese Field Epidemiology Training Program in Beijing
- Learnt about some public health projects initiated in Chaoyang District, Beijing, China

China-Japan Friendship Hospital, Beijing, China — Jun.2015-Jul.2015
Full-time Intern, Traditional Chinese Medicine Rheumatology Inpatient Department & Traditional Chinese Medicine Acupuncture Outpatient Department

- Wrote case reports
- Made rounds with doctors, attended professional trainings for interns, and auditing discussions on difficult cases
- Assisted the physician with acupotomy. including skin disinfection, and injection preparation, etc.

World Federation of Chinese Medicine Societies (WFCMS), Beijing, China — Sep.2014-Oct.2014
Full-time Intern, Qualification and Examination Division

- Collected examinees' information, made certificates and transcripts, and was involved in the exam paper review
- Participated in the preparation of The 11th World Congress of Chinese Medicine, including email translation, gift purchase and publicity video shooting, etc.

RESEARCH EXPERIENCE

Thesis Project: Polysubstance Use Patterns among Women with at-Risk Alcohol Consumption Dec.2017- present
Advisors: Bryan Lau, Associate Professor; Geetanjali Chander, Associate Professor
Bloomberg School of Public Health, Johns Hopkins University

Research Assistant at Oncospace (Radiation Oncology) Nov.2017-present
Supervisor: Todd McNutt, Associate Professor
School of Medicine, Johns Hopkins University

Course Projects
2017

- Inverse Probability Weighting and Multiple Imputation for Missing Data Problem: Application to Association between Homocysteine and Cognitive Function: the Baltimore Memory Study
- Height Trajectory in Children with Glomerular and Non-Glomerular Chronic Kidney Disease: the CKiD Study
- Research Proposal for Exercised-Induced Immune Marker Response after Spontaneous Intracerebral Hemorrhage

Bloomberg School of Public Health, Johns Hopkins University

Literature Research in Effectiveness of School-based Tobacco Control Programs in China Oct.2015-Nov.2015
Advisors: Gan Quan, Director; Liu Honge, Senior Technical Officer
the Union, China Office

Literature Research in Acupuncture Therapy's Efficacy on ADHD (attention deficit hyperactivity disorder)
Advisor: LI Xiaoli, Associate Professor May 2014-Jun.2014
Bejing University of Chinese Medicine

TEACHING/MENTOR EXPERIENCE

Teaching Assistant, Johns Hopkins University Sep. 2017-Oct.2017

- Course Title: Statistical Methods in Public Health I

Student Mentor, Johns Hopkins University Sep. 2017- present

PUBLICATION

- SHI, J.A., ZHAO, L. (2015). A Reflection on Hospital Bed Turnover Rate. *Chinese and Foreign Medical Research. 13*(22), 148-149. *(ISSN: 1674-6805)*

EXTRACURRICULUM ACTIVITIES

The Identity Clinic, Living Classrooms Foundation Jun.2017-present

- Helped ex-offenders and homeless people with the applications for official identifications, including birth certificate, social security card, and state ID

Trainee, Drama Performance Training Program, The Central Academy of Drama Feb.2016-Jun.2016

Planner, Volunteer, BUCM Physical and Mental Health for the Aged Promotion Project Nov.2014-Mar.2015

- Wrote a plan for the project of introducing the elder people to traditional Chinese medical health advice, organized volunteer groups and contact targeted communities
- Did survey about the health condition of the elder people in the neighborhood communities

Vice President of External Communication Department, BUCM Student Union Nov.2013-Nov.2014

- Proposed and prepared for the 2013 BUCM Culture Exchange Fair
- Organized the university volunteers for the 2014 Beijing and International Chinese Medicine Student Communication Camp

Broadcast, University Radio Station Oct.2011-Sep.2012

Member, University Arts Group
Oct.2011-Jul.2012

HONORS & AWARDS

- 2016 Beijing Higher Education Outstanding Graduates, Beijing, China 2016

- 2016 Outstanding Graduates of Beijing University of Chinese Medicine, Beijing, China 2016
- BUCM First-Class Scholarship (three years), Beijing, China 2011-2015
- BUCM "Three Good" Student (four years), Beijing, China 2011-2015
- Second Prize, The 3rd BLCU (Beijing Language and Culture University) International Translation and Interpreting Competition, News Editing and Translating (English-Chinese) Group, Beijing, China 2014
- BUCM Excellent League Member (two years), Beijing, China 2012-2014
- BUCM Excellent Student in Military Training, Beijing, China 2011

COMPUTER SKILLS

Microsoft Word, Excel, PowerPoint, Access, STATA, R, SAS, ArcGIS, RedCap.